

```
In [1]: import pandas as pd  
df = pd.read_csv("titanic_kaggle.csv")
```

```

In [30]: import matplotlib.pyplot as plt
import seaborn as sns
def find_outliers(column_name):

    sns.boxplot(data=df[column_name])
    plt.xlabel(column_name)
    plt.ylabel('Ticket price')
    plt.title(f'Boxplot of {column_name} on the Titanic')
    plt.show()

# Calculate the interquartile range (IQR)
Q1 = df[column_name].quantile(0.25)
Q3 = df[column_name].quantile(0.75)
IQR = Q3 - Q1
print("IQR:", IQR)

# Calculate the upper threshold
threshold = Q3 + (1.5 * IQR)
print("Threshold:", threshold)

# Identify outliers
outliers = df[column_name][df[column_name] > threshold]

plt.hist(outliers, bins=10)
plt.ylabel('No. of tickets')
plt.xlabel('Fare in Pounds')
plt.show()

# Create a cross-tabulation of survivors by gender
survivor_gender_counts = pd.crosstab(df['sex'], df['survived'])

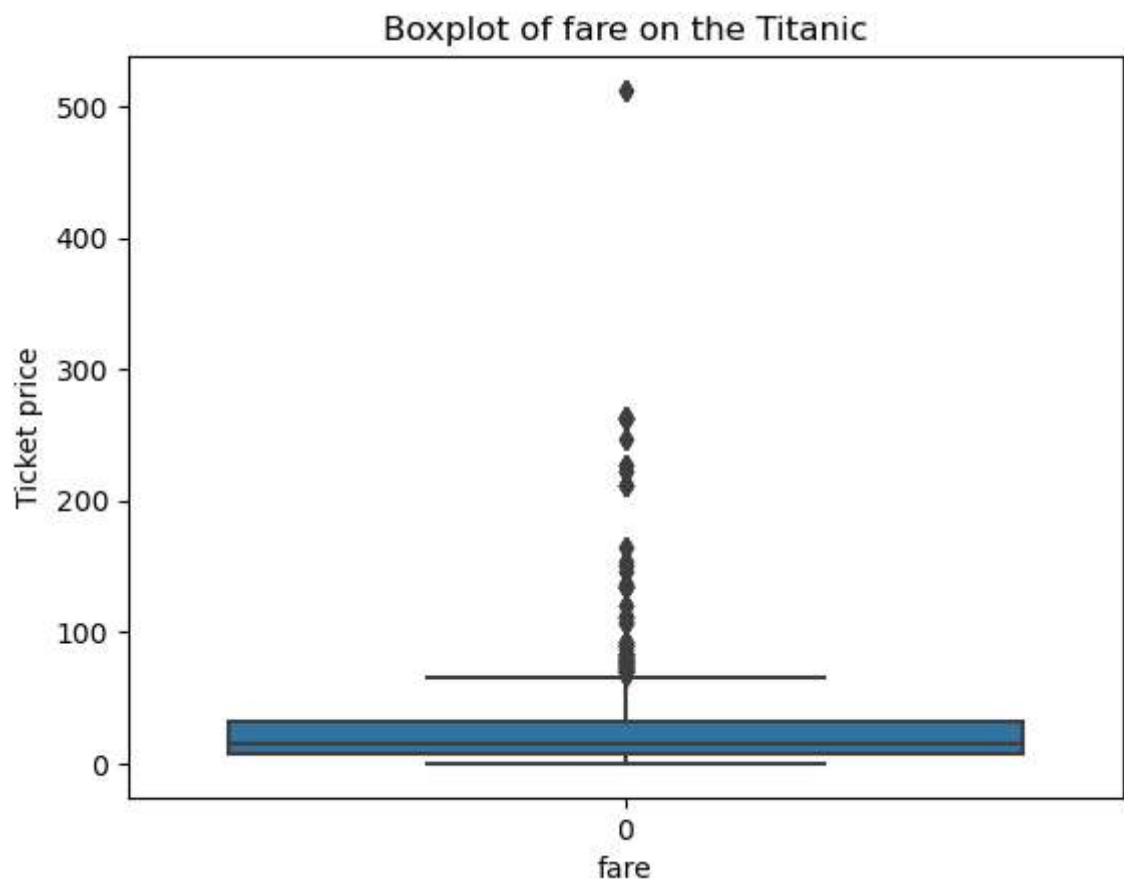
# Plot the stacked bar graph
survivor_gender_counts.plot(kind='bar', stacked=True)
plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Number of Survivors by Gender on the Titanic')
plt.legend(title='Survived', labels=['No', 'Yes'])
plt.show()

# Plot the bar graph
#sns.countplot(data=df[df['survived'] == 1], x='sex')
#plt.xlabel('Gender')
#plt.ylabel('Count')
#plt.title('Number of Survivors by Gender')
#plt.show()

# females_survived = df[(df['sex'] == 'female') & (df['survived'] == 1)]
# females_survived_count = females_survived.shape[0]
# print("Number of females who survived:", females_survived_count)
# males_survived = df[(df['sex'] == 'male') & (df['survived'] == 1)]
# males_survived_count = males_survived.shape[0]
# print("Number of males who survived:", males_survived_count)

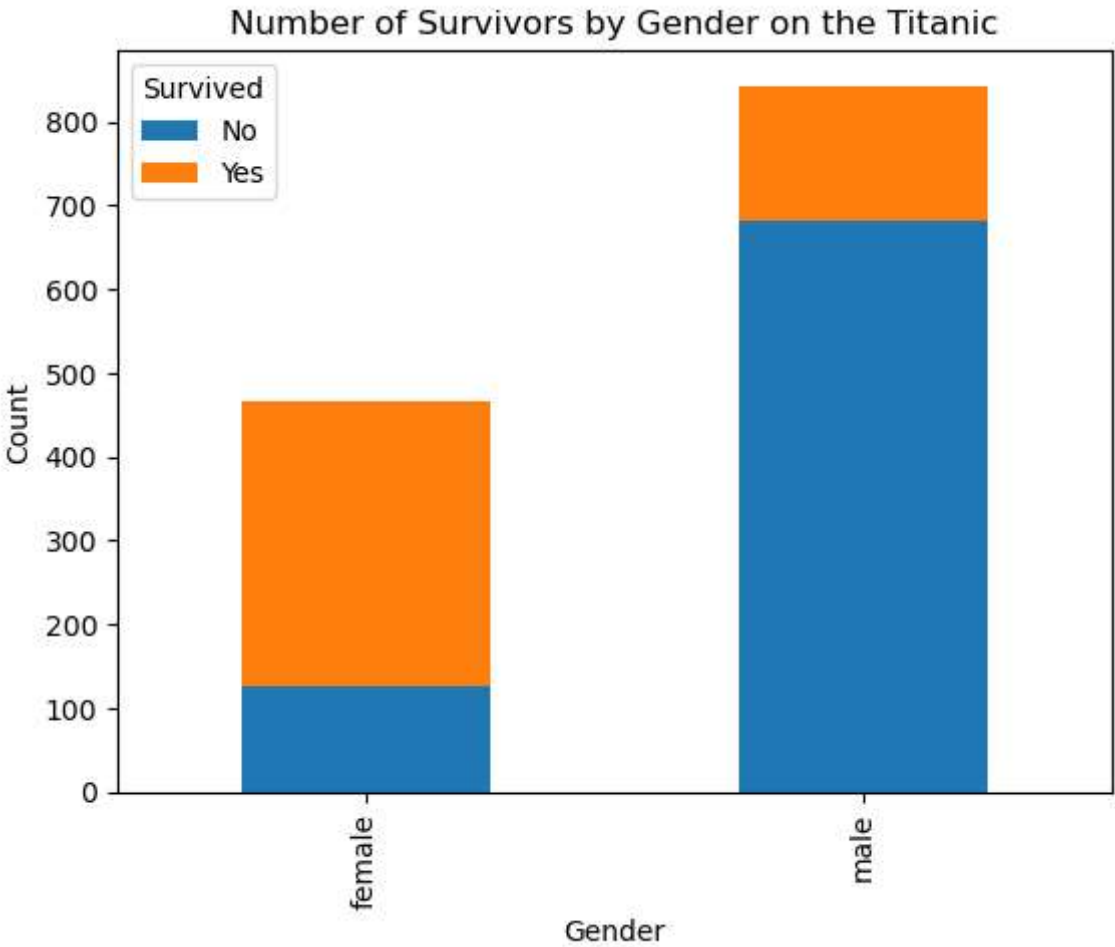
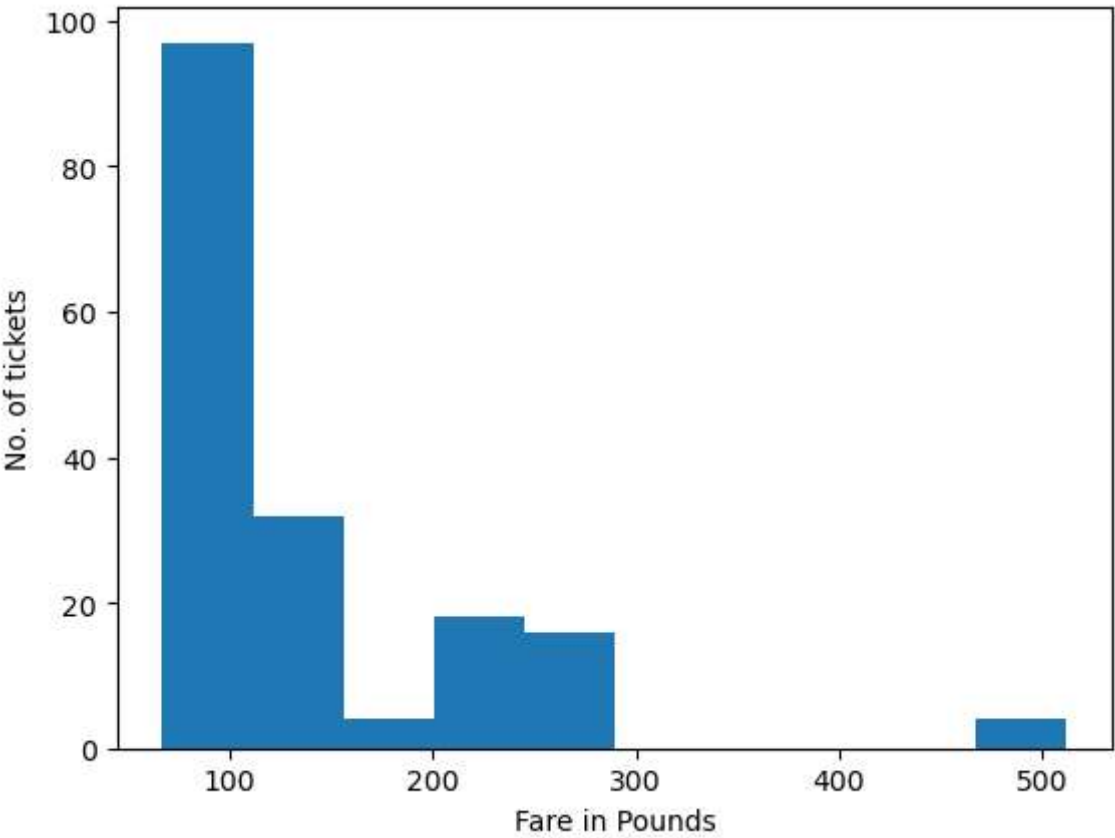
```

```
In [31]: find_outliers("fare")
```



IQR: 23.379199999999997

Threshold: 66.343799999999999



In [7]: `df.head()`

Out[7]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	S
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	S
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	S
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	S
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	S

In [9]: `females = df['sex'].value_counts()
females`

Out[9]: male 843
female 466
Name: sex, dtype: int64

In [12]: `df['survived'].value_counts()`

Out[12]: 0.0 809
1.0 500
Name: survived, dtype: int64

In [13]: `females_survived = df[(df['sex'] == 'female') & (df['survived'] == 1)]
females_survived_count = females_survived.shape[0]
print("Number of females who survived:", females_survived_count)`

Number of females who survived: 339

In [15]: `males_survived = df[(df['sex'] == 'male') & (df['survived'] == 1)]
males_survived_count = males_survived.shape[0]
print("Number of males who survived:", males_survived_count)`

Number of males who survived: 161

```

In [55]: import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.image as mpimg
def find_outliers(column_name):

    sns.boxplot(data=df[column_name])
    plt.xlabel(column_name)
    plt.ylabel('Ticket price')
    plt.title(f'Boxplot of {column_name} on the Titanic')
    plt.show()

# Calculate the interquartile range (IQR)
Q1 = df[column_name].quantile(0.25)
Q3 = df[column_name].quantile(0.75)
IQR = Q3 - Q1
print("IQR:", IQR)

# Calculate the upper threshold
threshold = Q3 + (1.5 * IQR)
print("Threshold:", threshold)

# Identify outliers
outliers = df[column_name][df[column_name] > threshold]

bg_image = mpimg.imread('C:/Users/AniruddhGurram/OneDrive - Vanguard Digital
# Create a figure with two subplots
fig, axes = plt.subplots(1, 2, figsize=(10, 5))
fig.figimage(bg_image, alpha=0.2, extend[0,1,0,1])
# Plot the histogram)
axes[0].hist(df[column_name], bins=10)
axes[0].set_ylabel('No. of tickets')
axes[0].set_xlabel('Fare in Pounds')
axes[0].set_title('Histogram of Fare on the Titanic')

# Create a cross-tabulation of survivors by gender
survivor_gender_counts = pd.crosstab(df['sex'], df['survived'])

# Plot the stacked bar graph
survivor_gender_counts.plot(kind='bar', stacked=True, ax=axes[1])
axes[1].set_xlabel('Gender')
axes[1].set_ylabel('Count')
axes[1].set_title('Number of Survivors by Gender on the Titanic')
axes[1].legend(title='Survived', labels=['No', 'Yes'])

# Adjust the layout and spacing
plt.tight_layout()

# Show the plot
plt.show()

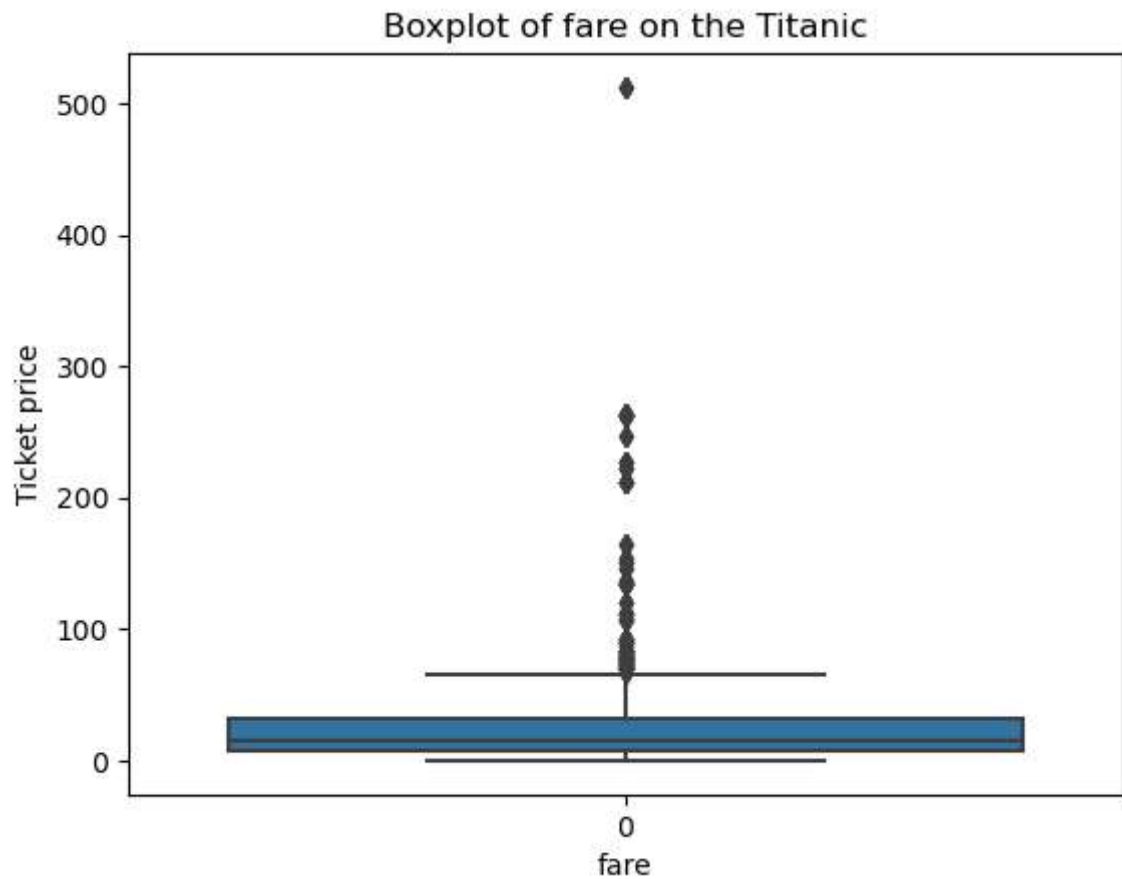
```

File "C:\Users\AniruddhGurram\AppData\Local\Temp\ipykernel_13156\332473964
3.py", line 25

fig.figimage(bg_image, alpha=0.2, extend[0,1,0,1])

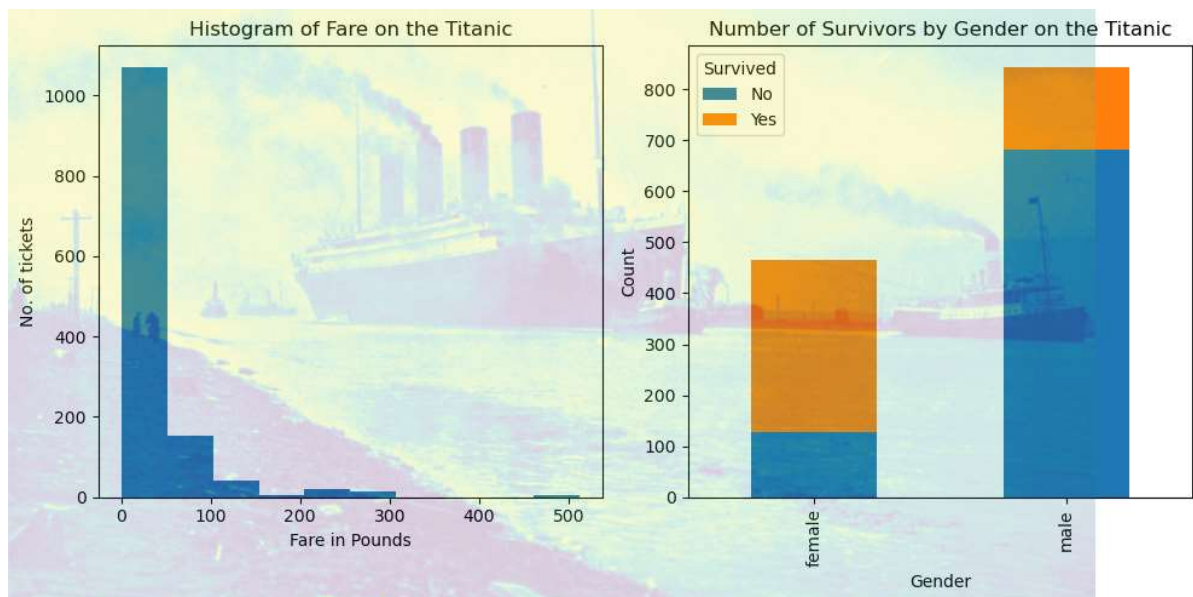
SyntaxError: positional argument follows keyword argument

```
In [56]: find_outliers("fare")
```



IQR: 23.379199999999997

Threshold: 66.34379999999999



```
In [ ]:
```

